Viewport-Aware Omnidirectional Video Streaming Using Visual Attention and Dynamic Tiles

PROBLEM AND OBJECTIVE

- Selected tiling scheme impacts the omnidirectional video (ODV) streaming efficiency.
- The most appropriate tiling scheme with the required encoding bitrate for each tile.
- A novel VA-driven quality metric for ODV.

ESTIMATION OF VISUAL ATTENTION MAP

- A viewpoint-based visual attention (VA) map for each chunk using the dataset in [1].
- Gaussian filtering is applied in the viewport domain.

EVALUATIONS

- Subjective experiments. A total of 17 participants (13 M and 4 F). Two groups: i) modelling of VA (12 par.); ii) validation of the proposed approach (5 par.).
- Benchmarks. Fixed-sized tiling schemes with the number of tiles $N = \{6,10,18\}$. PSNR measurement at the bitrate of the target DASH representations (2,5,10,15,20) Mbps.

OPTIMIZATION OF TILING SCHEME

Each given ODV consists of a set of tiles of various sizes and each tile, $t_i$, has $m$ different target encoding bitrates: $[R_1 \ldots R_m]$. Let $R_v$ be the target bitrate of the $v$-th DASH representation, for each $k$ chunk, the optimal tiling scheme $T_{sk}$ can be calculated as follows:

$$T_{s,k} = \max_{T \in T_k} Q_s(T, A_k, B(R_v, \cdot)), $$

where $Q_s(\cdot)$ represents the expected quality of $T$ applied to the content, taking into account the VA map $A_k$ and bitrate allocation function, $B(R_v, \cdot)$, subject to a target bitrate $R_v$.

Viewport-based bitrate allocation for tiles. The idea is to distribute the $R_v$ among tiles reinforcing those that are placed within a viewport with high VA values, $V_v$. Therefore, our proposed bitrate allocation depends on the target bitrate and the selected viewport: $B(R_v, V_v)$. The proposed method applies the viewport-based bitrate allocation method [2].

Estimation of the quality of a tiling scheme. The quality of a pair $(T_v, V_v)$ can be estimated as:

$$Q_v(T_v, A_k, B(V_v, \cdot)) = \sum_{i=1}^{N_v} P_A(V_v) Q_s(V_i, B(V_v, R_v)) + P_A(V_v) Q_s(V_i, B(V_v, R_v))$$

where $P_A(\cdot)$ is the probability that a viewport be watched, $Q_s(\cdot)$ represents the PSNR value, $O$ represents the remainder of the video area.

REFERENCES